



Compliance, ética e Inteligencia artificial
Claves para el desarrollo de una IA fiable

Iván Martínez López

imartinez@worldcomplianceassociation.com



www.worldcomplianceassociation.com

The ABC's of COMPLIANCE MONSTERS!



CORRUPCIÓN



SOBORNO



TRÁFICO DE INFLUENCIAS



COLUSIÓN



INTELIGENCIA ARTIFICIAL



MARCO REGULATORIO



FRAUDE INTERNO



TRIBUTARIO



PRÁCTICAS ANTICOMPETITIVAS



CONFLICTO DE INTERESES



PRIVACIDAD Y DATOS PERSONALES



CIBER SEGURIDAD



DERECHOS DE LOS CONSUMIDORES



EQUIDAD



RESPONSABILIDAD PENAL



SEGURIDAD



PROPIEDAD INTELECTUAL



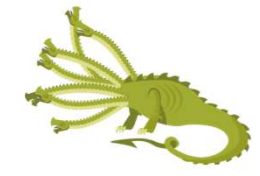
SOCIAL



DERECHOS LABORALES



LAVADO DE DINERO

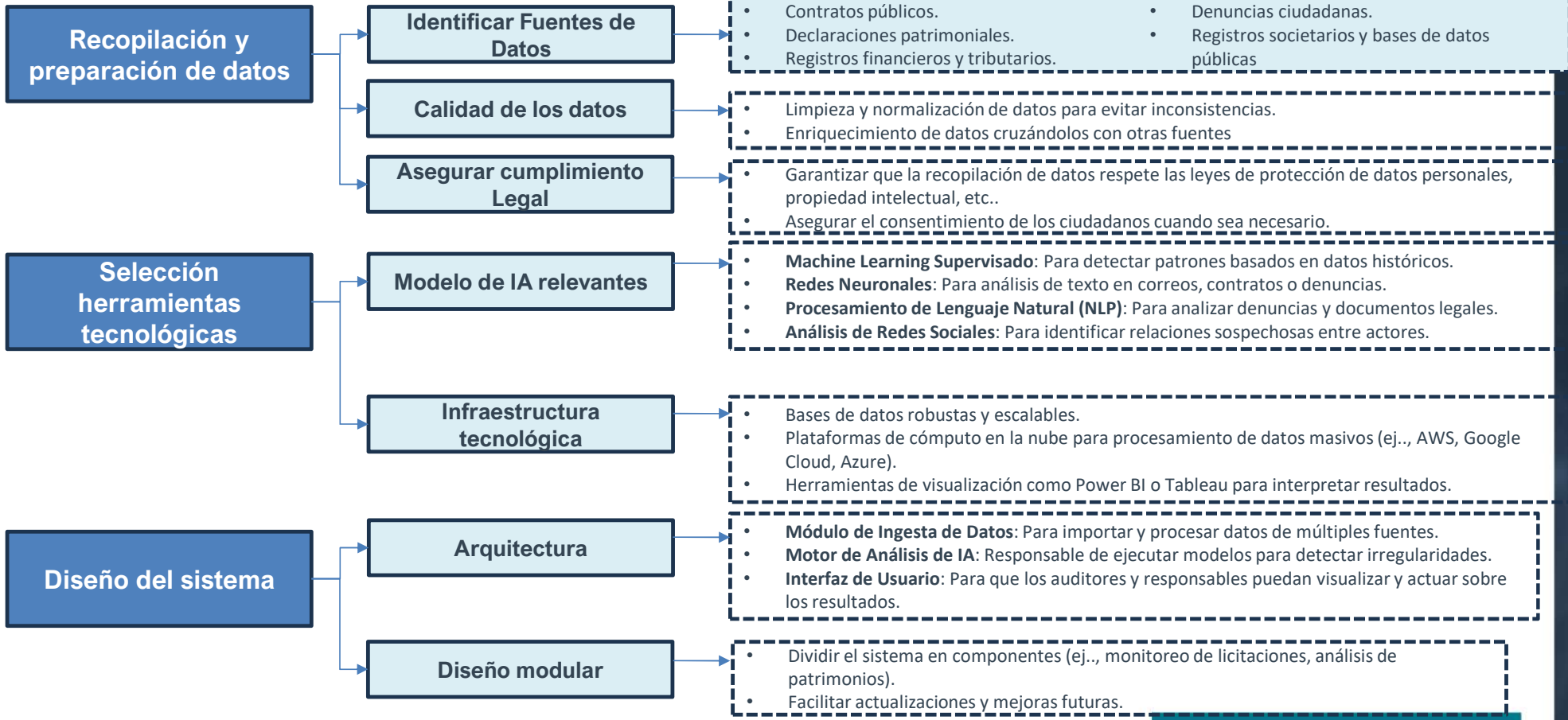


MEDIO AMBIENTE

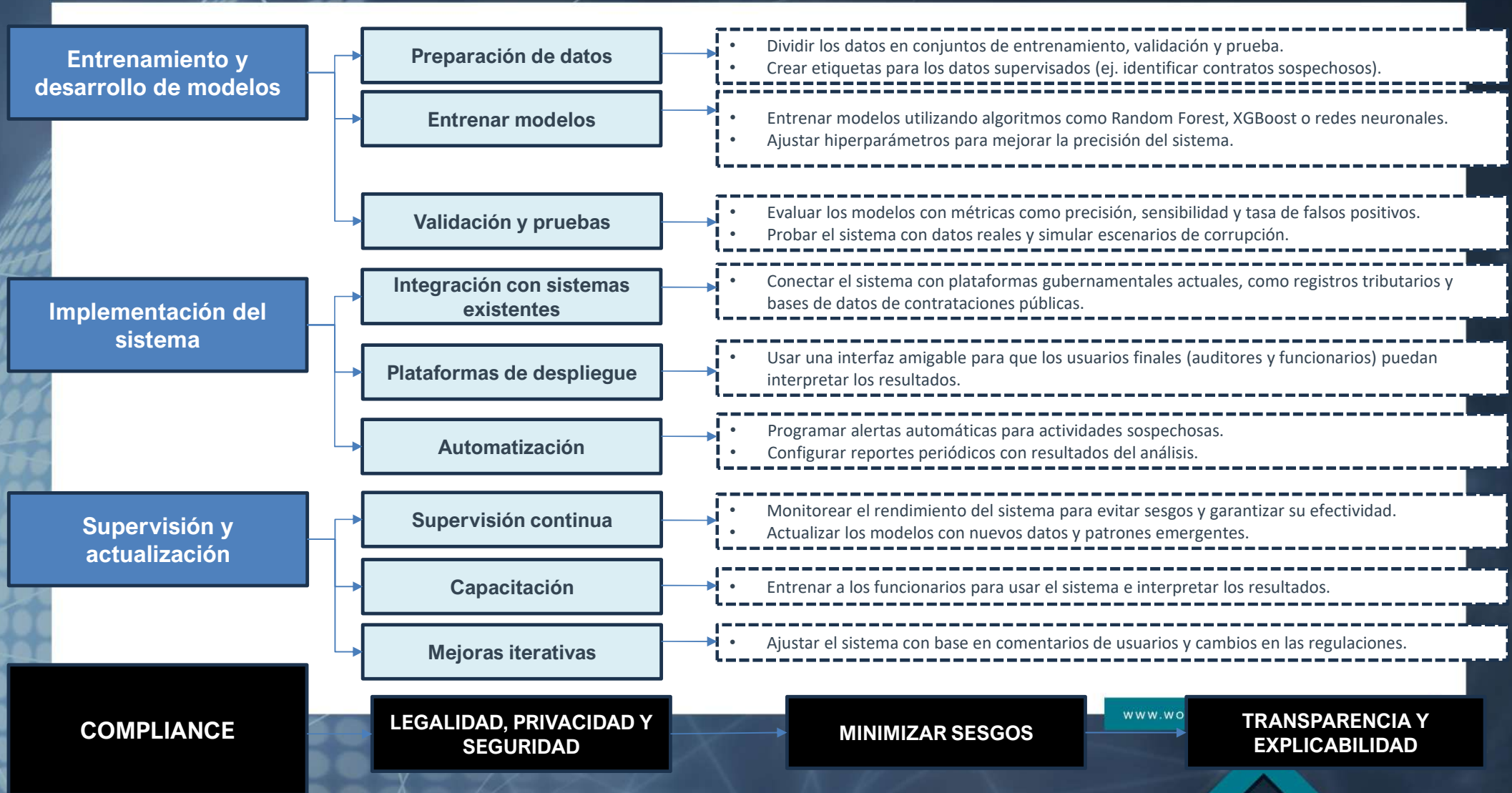
The ABC's of COMPLIANCE MONSTERS!

La
“Magia
de la
IA”





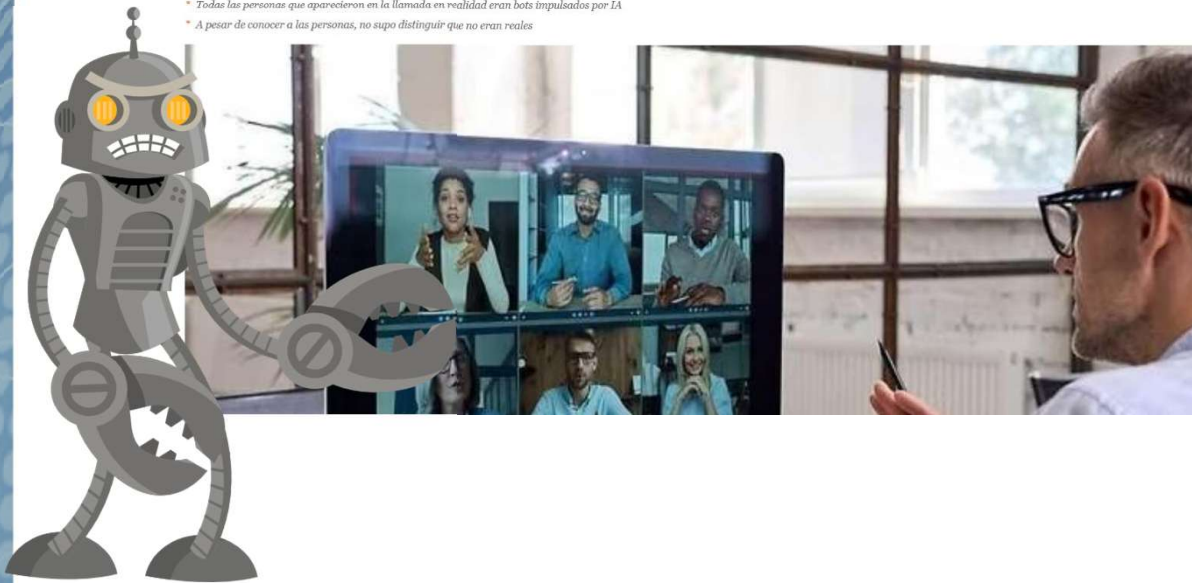
¿Sabemos como funciona la INTELIGENCIA ARTIFICIAL?



Riesgos emergentes... la Inteligencia Artificial

Estafa maestra con Inteligencia Artificial: suplantan al CEO y varios empleados en una videollamada para robar cientos de millones

* Todas las personas que aparecieron en la llamada en realidad eran bots impulsados por IA
* A pesar de conocer a las personas, no supo distinguir que no eran reales



¿Qué has hablado con ChatGPT? 100.000 conversaciones están ahora públicas en Google: de secretos de empresa a dramas de amores

- * ChatGPT ha confirmado que las conversaciones están en Google, pero de momento no ha podido deshacerlo
- * Pese a no ser una filtración, es una brecha de privacidad importante
- * Apple prepara su propio ChatGPT y quiere revolucionar la búsqueda en internet

Google ha cerrado el anuncio



Victor Millán

6/08/2023 - 11:56

¿Te imaginas que la conversación más íntima o más escabrosa que tuviste con ChatGPT acabe flotando en Google, accesible para cualquiera con un clic? Para miles de usuarios de ChatGPT, esa pesadilla ya es una realidad. **Entre los casi 100.000 chats expuestos hay desde borradores de contratos confidenciales hasta mensajes de amor no enviados**, un testimonio silencioso de cómo nos relacionamos con esta herramienta digital.

INTELIGENCIA ARTIFICIAL

Diella, la IA nombrada ministra en Albania para acabar con la corrupción en las contrataciones

V +

El primer ministro albanés, Edi Rama, asegura que el algoritmo garantizará que las inversiones públicas serán “perfectamente transparentes”



Diella es representada como una mujer con un vestido tradicional albanés (Gobierno de Alba

DESTACADOS Irán Israel Día de la Mujer Últimos videos Últimos audios

PANORAMA | GLOBAL

Actriz denuncia robo de su imagen para crear ministra de IA

Descubre DW 23/02/2020

Anila Bisha firmó un contrato limitado para una plataforma gubernamental, pero su rostro y voz terminaron personificando a Diella, presentada como la primera ministra de IA del mundo.

f x

Suscríbete por \$1020 Ingresar

Noticias hoy Dólar blue Guerra en Medio Oriente Javier Milei Tren Mtro Argentina Week Colapso en la F1 Líder de Irán Elecciones Colombia Homicidio hoy

En vivo Guerra entre Estados Unidos, Israel e Irán Javier Milei y sus medidas

The New York Times International Report

Albania creó un “Ministro de Inteligencia Artificial” para combatir la corrupción. Sus desarrolladores fueron acusados de corrupción

El avatar albanés conocido como Diella, un activista público contra la corrupción, ha sido descrito como el primer ministro del mundo creado por IA.

Imagen: Florion Gega/REUTERS

Diella es representada como una mujer con un vestido tradicional albanés (Gobierno de Alba



“ Compliance tendrá la
responsabilidad de asegurar
que los sistemas de IA
cumplan con estándares de
ética, transparencia,
legalidad, seguridad,
privacidad y explicabilidad.

”



¿Es mi sistema un "sistema de IA" según la Ley de IA de la UE?

<https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/>



EU AI Act

(Reglamento (UE) 2024/1689)

La Ley de IA de la UE es vista como un referente global en cuanto a la regulación de la inteligencia artificial. Establece un marco integral para **equilibrar la innovación tecnológica con la protección de los derechos de los ciudadanos**, posicionando a la UE como líder en la regulación ética de la IA. Además, se espera que inspire legislaciones similares en otras regiones del mundo.

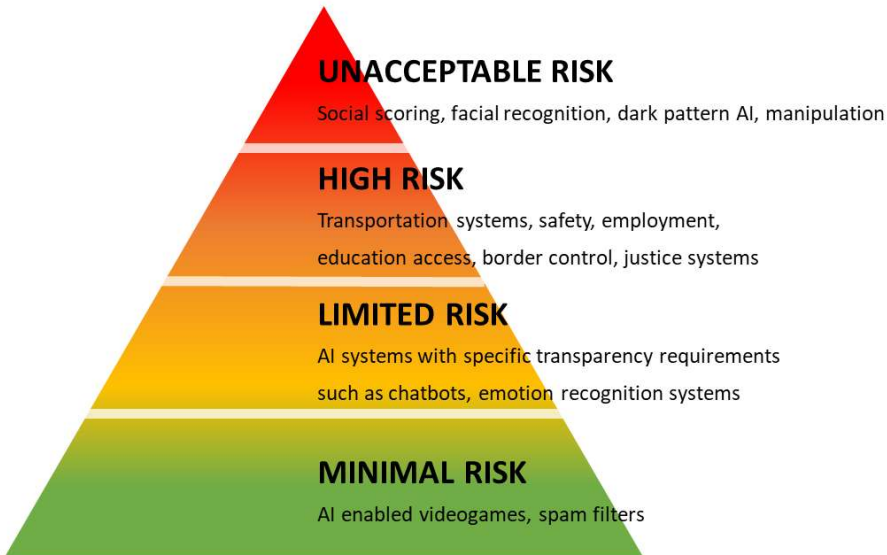
Este **enfoque preventivo y basado en riesgos** marca un hito importante en la relación entre la tecnología avanzada y la legislación, con el objetivo de construir un entorno seguro y confiable para el despliegue de IA.

El área de compliance relacionada con la inteligencia artificial (IA) está emergiendo como una de las más prometedoras a medida que entran en vigor regulaciones como la Ley de IA de la Unión Europea (AI Act) y otros marcos normativos a nivel global. Las oportunidades en materia de compliance para empresas e instituciones que operan en el campo de la IA son variadas y crecientes, especialmente debido a la necesidad de asegurar que los **sistemas de IA cumplan con estándares éticos, de transparencia, privacidad y de seguridad**.

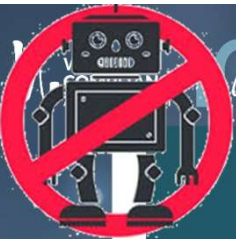
Un sistema de inteligencia artificial (sistema IA) se define como: Un sistema basado en máquinas diseñado para operar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras su despliegue y que, para objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales.

EU AI Act

(Reglamento (UE) 2024/1689)



- ➔ **Riesgo inaceptable:** Se prohíben los sistemas de IA que presenten un riesgo significativo para los derechos fundamentales y la seguridad de las personas. Ejemplos incluyen el uso de IA para manipular el comportamiento humano o sistemas de vigilancia masiva con IA.
- ➔ **Alto riesgo:** Sistemas de IA que impactan áreas críticas como infraestructuras, educación, empleo, servicios financieros, y acceso a la justicia. Estos sistemas estarán sujetos a estrictos controles, auditorías, y requisitos de transparencia.
- ➔ **Riesgo limitado:** Sistemas de IA que interactúan con los usuarios, como chatbots. Estos deben cumplir con ciertas obligaciones de transparencia para que los usuarios sepan que están interactuando con una IA.
- ➔ **Riesgo mínimo o nulo:** Aplicaciones de IA con bajo impacto, como los filtros de spam. No están sujetas a regulación estricta, pero deben seguir ciertas buenas prácticas.



Years
ANNIVERSARY
CELEBRATION

SISTEMAS PROHIBIDOS

EU AI Act

(Reglamento (UE) 2024/1689)



Despliegue de técnicas subliminales o manipuladoras



Explotación de vulnerabilidades (edad, discapacidad, situación económica)



Categorización biométrica para inferir atributos sensibles



Puntuación social



Evaluación de riesgo criminal basada solo en perfiles



Creación de bases de datos faciales con scraping masivo



Detección de emociones en entorno laboral o educativo (salvo seguridad o salud)



Identificación biométrica remota en tiempo real en espacios públicos

- **Despliegue de técnicas subliminales, manipuladoras o engañosas** para distorsionar el comportamiento y perjudicar la toma de decisiones con conocimiento de causa, causando un daño significativo.
- **Explotar las vulnerabilidades** relacionadas con la edad, la discapacidad o las circunstancias socioeconómicas para distorsionar el comportamiento, causando daños significativos.
- **Sistemas de categorización biométrica que infieran atributos sensibles** (raza, opiniones políticas, afiliación sindical, creencias religiosas o filosóficas, vida sexual u orientación sexual), excepto el etiquetado o filtrado de conjuntos de datos biométricos adquiridos legalmente o cuando las fuerzas de seguridad categoricen datos biométricos.
- **Puntuación social**, es decir, evaluar o clasificar a individuos o grupos basándose en comportamientos sociales o rasgos personales, causando un trato perjudicial o desfavorable a esas personas.
- **Evaluar el riesgo de que un individuo cometa delitos penales** basándose únicamente en perfiles o rasgos de personalidad, excepto cuando se utilice para aumentar las evaluaciones humanas basadas en hechos objetivos y verificables directamente relacionados con la actividad delictiva.
- **Compilación de bases de datos de reconocimiento facial** mediante el raspado no selectivo de imágenes faciales de Internet o de grabaciones de CCTV.
- **Inferir emociones en lugares de trabajo o centros educativos**, salvo por razones médicas o de seguridad.
- **Identificación biométrica remota (RBI)** "en tiempo real" en espacios de acceso público para las fuerzas de seguridad excepto: búsqueda de personas desaparecidas, víctimas de secuestros y personas víctimas de la trata de seres humanos o la explotación sexual; prevenir una amenaza sustancial e inminente para la vida, o un ataque terrorista previsible; o identificar a sospechosos de delitos graves (por ejemplo, asesinato, violación, robo a mano armada, tráfico de estupefacientes y armas ilegales, delincuencia organizada y delitos contra el medio ambiente, etc.).

Sistemas de ALTO RIESGO

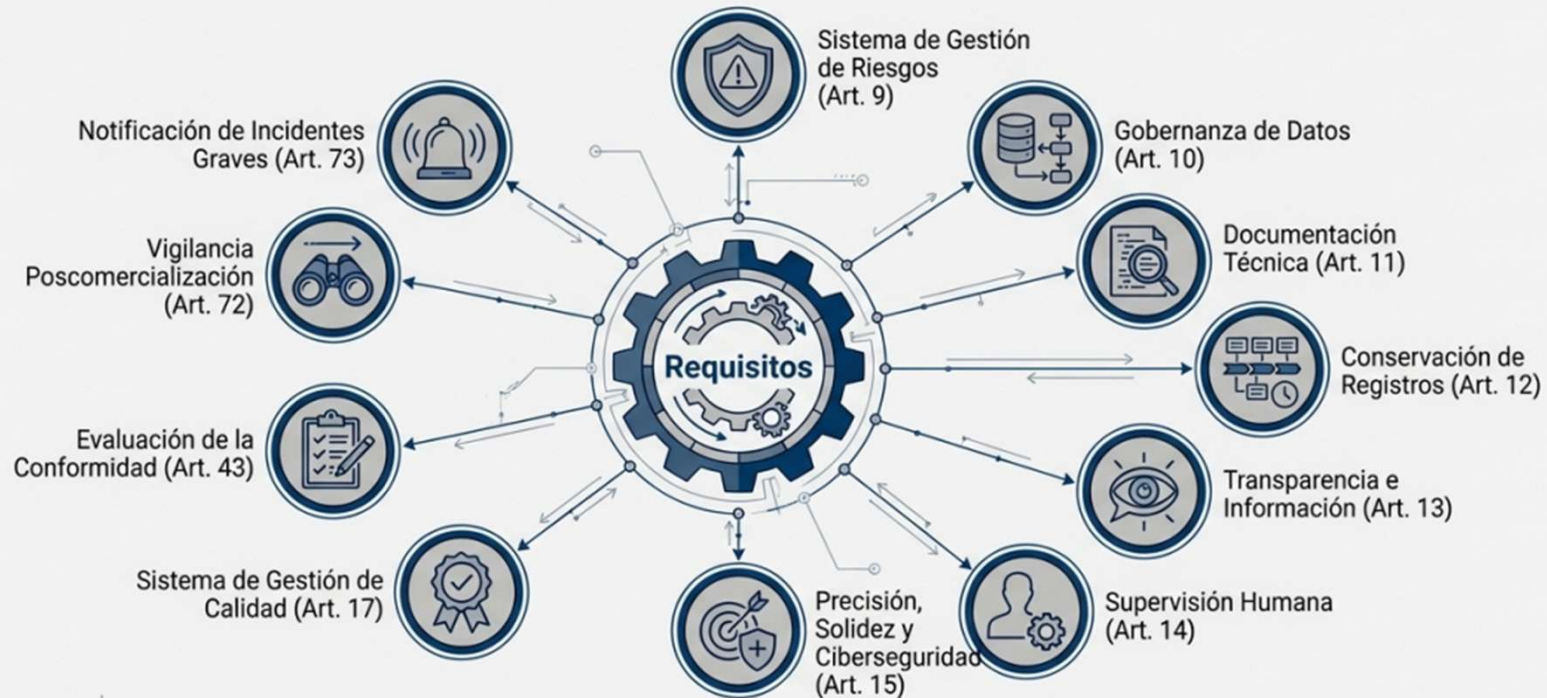


Fuente: AESIA

Sistemas de ALTO RIESGO

Obligaciones Fundamentales para Sistemas de Alto Riesgo

Para garantizar la seguridad, fiabilidad y equidad, los proveedores de sistemas de IA de alto riesgo deben cumplir con un conjunto de requisitos rigurosos a lo largo de todo el ciclo de vida del sistema.



Cada una de estas obligaciones está respaldada por guías técnicas detalladas para una implementación práctica.

Fuente: AESIA

| Tipo de sistema | Fecha de introducción/puesta en servicio | Cumplimiento de Obligaciones |
|--|--|------------------------------|
| Sistemas Prohibidos | N/A | Febrero 2025 |
| Sistemas de Alto Riesgo (Anexo III) | Después del 2 de agosto de 2026 | Diciembre 2027 (aprox.) |
| Sistemas de Alto Riesgo (Utilizados por Autoridades Públicas) | Antes del 2 de agosto de 2026 | Agosto 2030 |
| GPAI (Modelos de IA de Uso General) | Después del 2 de agosto de 2025 | Agosto 2026 |
| Obligaciones de Transparencia (Art. 50) | Después del 2 de agosto de 2026 | Febrero 2027 (aprox.) |

Fuente: AESIA

Requisitos Clave: Supervisión Humana y Transparencia (Artículos 13 y 14)

Objetivo del Requisito: Garantizar que los sistemas de IA de alto riesgo puedan ser comprendidos, utilizados adecuadamente y supervisados de manera efectiva por personas para prevenir o minimizar riesgos.

Pilar 1: Transparencia (Art. 13)

- **Diseño**
El sistema debe ser diseñado para que los responsables del despliegue puedan interpretar sus resultados y utilizarlo adecuadamente.
- **Instrucciones de Uso**
Deben incluir información concisa, completa y clara sobre:
 - Identidad y contacto del proveedor.
 - Características, capacidades y limitaciones del sistema.
 - Niveles de precisión, solidez y ciberseguridad.
 - Finalidad prevista y cualquier uso indebido razonablemente previsible.



Pilar 2: Supervisión Humana (Art. 14)

- **Objetivo**
Permitir que las personas vigilen eficazmente el funcionamiento del sistema y puedan intervenir.
- **Medidas de Diseño**
El sistema debe:
 - Ser comprensible para la persona supervisora.
 - Permitir la anulación, la detención o la interrupción segura del sistema.
 - Minimizar riesgos como el sesgo de automatización.
- **Interfaces**
Las interfaces humano-máquina deben ser adecuadas para permitir una supervisión informada y consciente.

Fuente: AESIA

SANCIONES POR INCUMPLIMIENTO DEL AI ACT



MULTA MÁXIMA
Hasta **35 MILL€**
o **7% del negocio global anual**

⚠️ INFRACCIONES MUY GRAVES

- **Uso de IA prohibida** (riesgo inaceptable)
 - Manipulación subliminal que afecte al comportamiento
 - Explotación de vulnerabilidades de colectivos
 - Social scoring (calificación social)
 - Identificación biométrica remota legal (excepto en contadas excepciones)

⚠️ INFRACCIONES GRAVES

- **Incumplimiento de normas del AI Act** (ej: IA de alto riesgo)
 - Gestión de riesgos y gobernanza de datos
 - Garantía de calidad de datos
 - Documentación técnica
 - Trazabilidad y supervisión humana
 - Ciberseguridad

Gestión de la IA en organizaciones

- ➔ **Requisitos y orientación** para establecer, implementar, mantener y mejorar continuamente un sistema de gestión de IA (inteligencia artificial) dentro del contexto de una organización.
- ➔ **Destinado a** ser utilizado por una organización que proporciona o utiliza productos o servicios que utilizan sistemas de IA manera responsable para lograr sus objetivos y cumplir con los requisitos aplicables.



Estructura de Alto Nivel

+

Norma Tipo A

+

ANEXOS

INTERNATIONAL
STANDARD

ISO/IEC
42001

First edition
2023-12

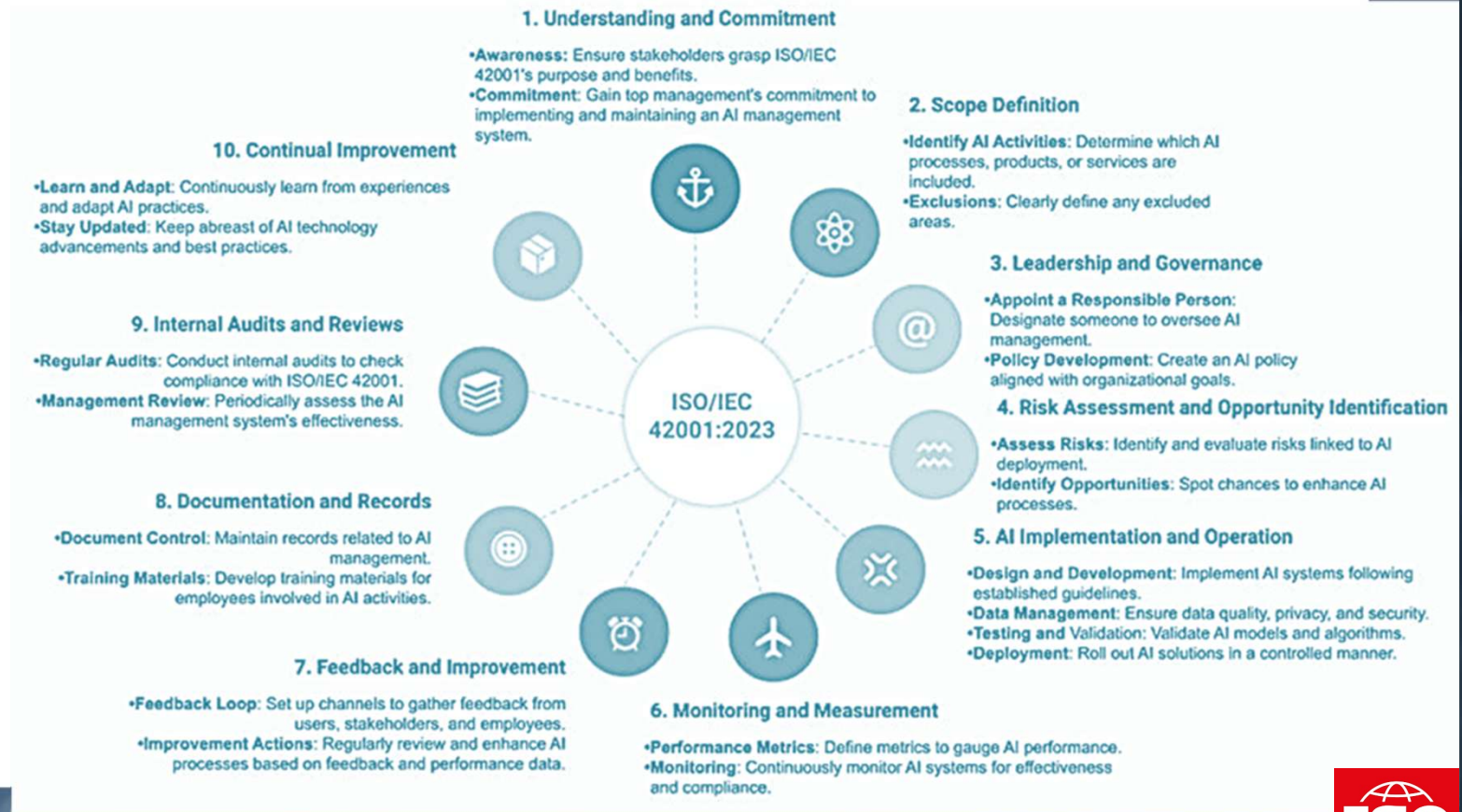
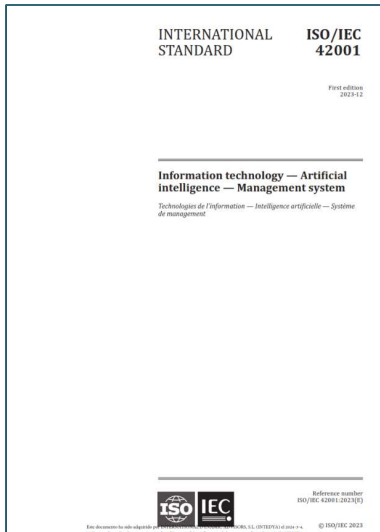
Information technology — Artificial
intelligence — Management system

Technologies de l'information — Intelligence artificielle — Système
de management



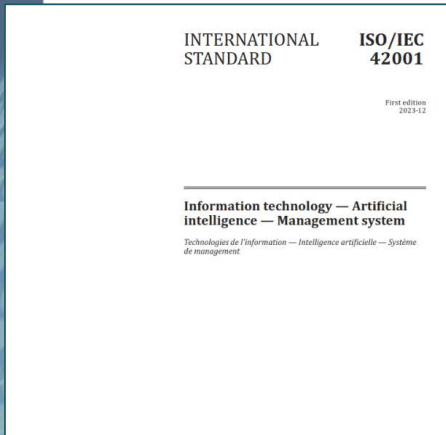
Reference number
ISO/IEC 42001:2023(E)

Gestión de la IA en organizaciones



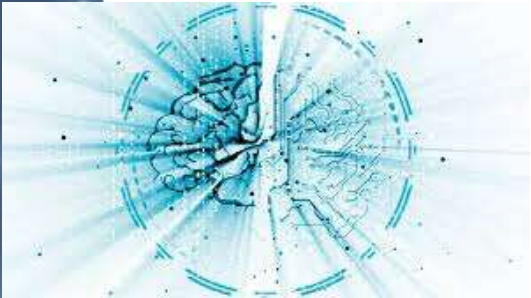
Gestión de la IA en organizaciones

Controles específicos



- ANEXO A: Objetivos de control y controles
- ANEXOS B: Guía de implementación de controles
- ANEXOS C: Posibles objetivos organizacionales relacionados con la IA y fuentes de riesgo
- ANEXO D: Uso del sistema de gestión de IA en todos los dominios o sectores





ISO 23894 Gestión de Riesgos en materia de inteligencia artificial



Proporciona orientación sobre cómo las organizaciones que desarrollan productos, implementan o utilizan productos, sistemas y servicios que utilizan inteligencia artificial (IA) pueden **gestionar los riesgos específicamente relacionados con la IA.**



La guía también tiene como objetivo ayudar a las organizaciones a integrar la gestión de riesgos en sus actividades y funciones relacionadas con la IA. Además, describe los procesos para la implementación e **integración efectiva de la gestión de riesgos de la IA.**

INTERNATIONAL STANDARD ISO/IEC 23894

First edition
2023-02

Information technology — Artificial intelligence — Guidance on risk management

*Technologies de l'information — Intelligence artificielle —
Recommandations relatives au management du risque*



Reference number
ISO/IEC 23894:2023(E)

Anexo A - ISO/IEC 23894:2023

(Objetivos relacionados con los riesgos de IA)

Objetivo

A.2 RESPONSABILIDAD (ACCOUNTABILITY)



Definir quién es responsable de las decisiones y efectos de la IA (organización vs sistema).

IA toma decisiones automatizadas sin un responsable claro; auditor no puede asignar responsabilidades.

A.3 COMPETENCIA Y EXPERIENCIA EN IA



Asegurar que quienes desarrollan, usan o supervisan IA tienen conocimientos adecuados.

Proyecto gestionado por personal sin habilidades en datos o IA: errores no detectados.

A.4 DATOS DE ENTRENAMIENTO Y PRUEBA



Los datos deben ser suficientes, representativos, actuales y de calidad.

Datos incompletos o sesgados: discriminación en decisiones (ej. préstamos, selección laboral).

A.5 IMPACTO AMBIENTAL



Considerar consumo energético, huella de carbono y sostenibilidad.

Entrenamiento de modelos generativos consume energía excesiva sin compensación ambiental.

A.6 EQUIDAD (FAIRNESS)



Evitar sesgos y resultados discriminatorios.

Algoritmo de contratación excluye a mujeres o minorías por datos históricos sesgados.

A.7 MANTENIBILIDAD



Posibilidad de actualizar, corregir y mejorar la IA sin perder control o trazabilidad.

Modelo entrenado no puede actualizarse porque no se documentaron versiones previas.

¿Qué implica?

Ejemplos de Riesgo Asociado

Anexo A - ISO/IEC 23894:2023

(Objetivos relacionados con los riesgos de IA)

Objetivo

¿Qué implica?

Ejemplos de Riesgo Asociado

A.8

PRIVACIDAD



Protección de datos personales y cumplimiento de leyes como GDPR.

IA infiere datos sensibles (religión, salud) sin consentimiento.

A.9

ROBUSTEZ



Capacidad de mantener su función ante errores, datos inesperados o ataques adversarios.

Un modelo de visión identifica señales erróneas con simples pegatinas adversarias.

A.10

SEGURIDAD (SAFETY)



Evitar daños a personas, bienes o el entorno.

Robot industrial controlado por IA ejecuta un movimiento peligroso sin freno humano.

A.11

CIBERSEGURIDAD



Proteger el sistema contra ataques, robo de modelos, manipulación de datos.

Ataque de "model stealing" o envenenamiento de datos en un sistema médico.

A.12

TRANSPARENCIA Y EXPLICABILIDAD



Posibilidad de entender cómo funciona, qué datos usa y cómo decide la IA.

IA bancaria rechaza crédito, pero no puede explicarse el motivo: falta de confianza y cumplimiento legal.

Fuentes de riesgo en IA Anexo B ISO/IEC 23894

B.2

COMPLEJIDAD DEL ENTORNO



La IA opera en entornos dinámicos o impredecibles, lo que aumenta incertidumbre y errores.

- Vehículo autónomo en clima extremo o carreteras no previstas.
- IA médica enfrentando enfermedades raras no incluidas en datos de entrenamiento.

B.3

FALTA DE TRANSPARENCIA Y EXPLICABILIDAD



Decisiones de la IA difíciles de entender o justificar para humanos.

- Modelo que rechaza una hipoteca y no se puede explicar por qué.
- "Caja negra" en redes neuronales profundas.

B.4

NIVEL DE AUTOMATIZACIÓN



Riesgos derivados del grado de autonomía y del traspaso de control humano-máquina.

- Piloto automático que solicita intervención humana demasiado tarde.
- Operarios desconectados por exceso de confianza en la IA.

B.5

APRENDIZAJE AUTOMÁTICO Y DATOS



Riesgos ligados a calidad de datos, sesgos, ataques o cambios durante el aprendizaje.

- Datos de entrenamiento sesgados que generan discriminación.
- Ataques adversarios (data poisoning, adversarial examples).
- Sistemas que siguen aprendiendo y modifican su comportamiento sin control.

B.6

HARDWARE Y FALLOS TÉCNICOS



Errores o malfuncionamiento en componentes físicos o infraestructura del sistema.

- Sensor LIDAR sucio o desalineado.
- Fallos de GPU/CPU durante el procesamiento.
- Pérdida de conectividad en sistemas en la nube.

B.7

DEPENDENCIA DE TERCEROS / ECOSISTEMA



Uso de proveedores, datos o modelos externos que introducen riesgos externos.

- API de proveedor externo deja de funcionar.
- Modelo reentrenado con licencias dudosas o datos ilegales.
- Falta de soporte o actualizaciones de software utilizado

B.8

MADUREZ TECNOLÓGICA



Uso de tecnologías de IA poco maduras o demasiado maduras (deuda técnica).

- Algoritmo experimental sin pruebas suficientes.
- Uso de IA antigua que ya no recibe parches de seguridad.
- Soluciones "beta" aplicadas en sectores críticos como salud o justicia.

What is the AI Risk Repository?

<https://airisk.mit.edu/>



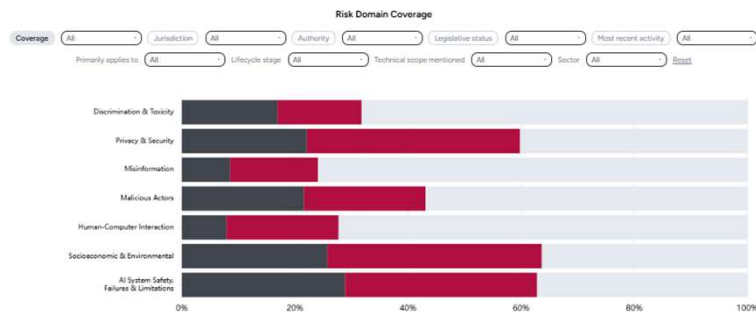
The AI Risk Repository has three parts:

- The AI Risk Database captures 1700+ risks extracted from 74 existing frameworks and classifications of AI risks
- The Causal Taxonomy of AI Risks classifies how, when, and why these risks occur
- The Domain Taxonomy of AI Risks classifies these risks into 7 domains and 24 subdomains (e.g., “False or misleading information”)

The repository is part of the MIT AI Risk Initiative, which aims to increase awareness and adoption of best practice AI risk management across the AI ecosystem.

| | | |
|---|--|--|
| 1. Discrimination & Toxicity 1.1 Unfair discrimination and misrepresentation 1.2 Exposure to toxic content 1.3 Unequal performance across groups | 2. Privacy & Security 2.1 Compromise of privacy by obtaining, leaking or correctly inferring sensitive information 2.2 AI system security vulnerabilities and attacks | 3. Misinformation 3.1 False or misleading information 3.2 Pollution of information ecosystem and loss of consensus reality |
| 4. Malicious Actors 4.1 Disinformation, surveillance, and influence at scale 4.2 Fraud, scams, and targeted manipulation 4.3 Cyberattacks, weapons development or use and mass harm | 5. Human-Computer Interaction 5.1 Overreliance and unsafe use 5.2 Loss of human agency and autonomy | 6. Socioeconomic & Environmental 6.1 Power centralization and unfair distribution of benefits 6.2 Increased inequality and decline in employment quality 6.3 Economic and cultural devaluation of human effort 6.4 Competitive dynamics 6.5 Governance failure 6.6 Environmental harm |
| 7. AI System Safety, Failures, & Limitations 7.1 AI pursuing its own goals in conflict with human goals or values 7.2 AI possessing dangerous capabilities 7.3 Lack of capability or robustness 7.4 Lack of transparency or interpretability 7.5 AI welfare and rights 7.6 Multi-agent risks | | |

What AI Risks Are Currently Being Governed?



ISO 42005:2025

Evaluación de Impacto del sistema de IA

- 1 Definir el alcance
- 2 Asignar responsabilidades
- 3 Recolección y análisis de información
- 4 Identificar impactos y riesgos
- 5 Evaluar nivel de impacto
- 6 Establecer medidas de mitigación
- 7 Registro y aprobación
- 8 Seguimiento

Proceso formal, documentado y basado en evidencia mediante el cual una organización **identifica, evalúa y gestiona los impactos reales o previsibles que un sistema de IA puede generar sobre personas, grupos, sociedad, derechos, entorno y otros sistemas.**



FINAL DRAFT
International
Standard

ISO/IEC FDIS
42005

Information technology — Artificial
intelligence — AI system impact
assessment

ISO/IEC JTC 1/SC 42

Secretariat: ANSI

Voting begins on:

2025-02-19

Voting terminates on:

2025-04-16

RECIPIENTS OF THIS DRAFT ARE INVITED TO SUBMIT
WITH THEIR COMMENTS, NOTIFICATION OF ANY
RELEVANT PATENT RIGHTS OR RIGHTS THAT ARE AWARE
AND TO PROVIDE SUPPORTING DOCUMENTATION.
IN ADDITION TO THEIR EVALUATION AS
BEING ACCEPTABLE FOR INDUSTRIAL, TECHNICAL,
SOCIAL, COMMERCIAL AND USER PURPOSES, DRAFT
INTERNATIONAL STANDARDS MAY ON OCCASION HAVE
TO BECOME CHANGED TO WHICH REFERENCE MAY BE
MADE IN NATIONAL REGULATIONS.

Reference number
ISO/IEC FDIS 42005:2025(en)

© ISO/IEC 2025

!!!ALERTA SPOILER!!!



DOSSIER DE CERTIFICACIÓN PROFESIONAL

WORLD COMPLIANCE ASSOCIATION **10** Years ANNIVERSARY CELEBRATION



PROFESIONAL CERTIFICATION
AI
ARTIFICIAL INTELLIGENCE OFFICER
WORLD COMPLIANCE ASSOCIATION

CERTIFICACIÓN PROFESIONAL
Artificial Intelligence Officer (AI Officer)

www.worldcomplianceassociation.com



DISCLAIMER



Ningún Robot ha sufrido daños
— durante la realización de esta presentación —





Compliance, ética e Inteligencia artificial

Claves para el desarrollo de una IA fiable

Iván Martínez López

imartinez@worldcomplianceassociation.com



www.worldcomplianceassociation.com